ЛЕКЦИЯ 3
3 модуль

МОДЕЛИ БИНАРНОГО ВЫБОРА

Демидова О.А.

План лекции

1) **Модель линейной вероятности, ее недостатки**
2) **Логит и Пробит модели, их оценивание**
3) **Интерпретация результатов оценивания моделей с бинарными зависимыми переменными**
4) **Показатели качества оценки моделей бинарного выбора**

Модель линейной вероятности

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

$$Y_i = E(Y_i) + \varepsilon_i$$

Найдем математическое ожидание Y_i.

$$p_i = p(Y_i = 1)$$

$$E(Y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i =$$

$$= \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

$$p_i = p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

Модель линейной вероятности

$$p_i = p(Y_i = 1) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$$

Если мы будем оценивать модель с качественной зависимой переменной, как и ранее, с помощью МНК, мы получим указанную выше модель, называемую моделью линейной вероятности.

Модель линейной вероятности

Однако линейная вероятностная модель имеет ряд серьезных недостатков:

Одним из главных недостатков линейной вероятностной модели является следующий : оцененные значения вероятности могут оказаться больше 1 или меньше 0,

Распределение случайного члена не является нормальным,

Можно показать, что дисперсия случайного члена зависит от X, таким образом, имеет место проблема гетероскедастичности.

Модели бинарного выбора

Обычным способом решения этой проблемы является предположение о том, что вероятность является S – образной функцией от переменной *Z*, *F*(*Z*) принимает значения на интервале (0, 1), где *Z* является линейной функцией от объясняющих переменных.

Модели бинарного выбора

Другой способ получения модели:

$P(Y_i = 1) = F(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})$ (*)

Предположим, что существует латентная переменная Y_i^*,

связанная с переменной X обычным регрессионным уравнением:

$Y_i^* = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$, $i = 1,\ldots,n$,

где возмущения ε_i независимы и одинаково распределены,

$E(\varepsilon_i) = 0$, $D(\varepsilon_i) = \sigma^2$

и F – функция распределения нормированных возмущений.

Функция плотности нормированных возмущений симметрична.

Модели бинарного выбора

Y_i^* - латентная (ненаблюдаемая переменная)

$Y_i = 1$, если $Y_i^* \geq 0$, $i = 1,\ldots,n$,

$Y_i = 0$, если $Y_i^* < 0$, $i = 1,\ldots,n$,

Модели бинарного выбора

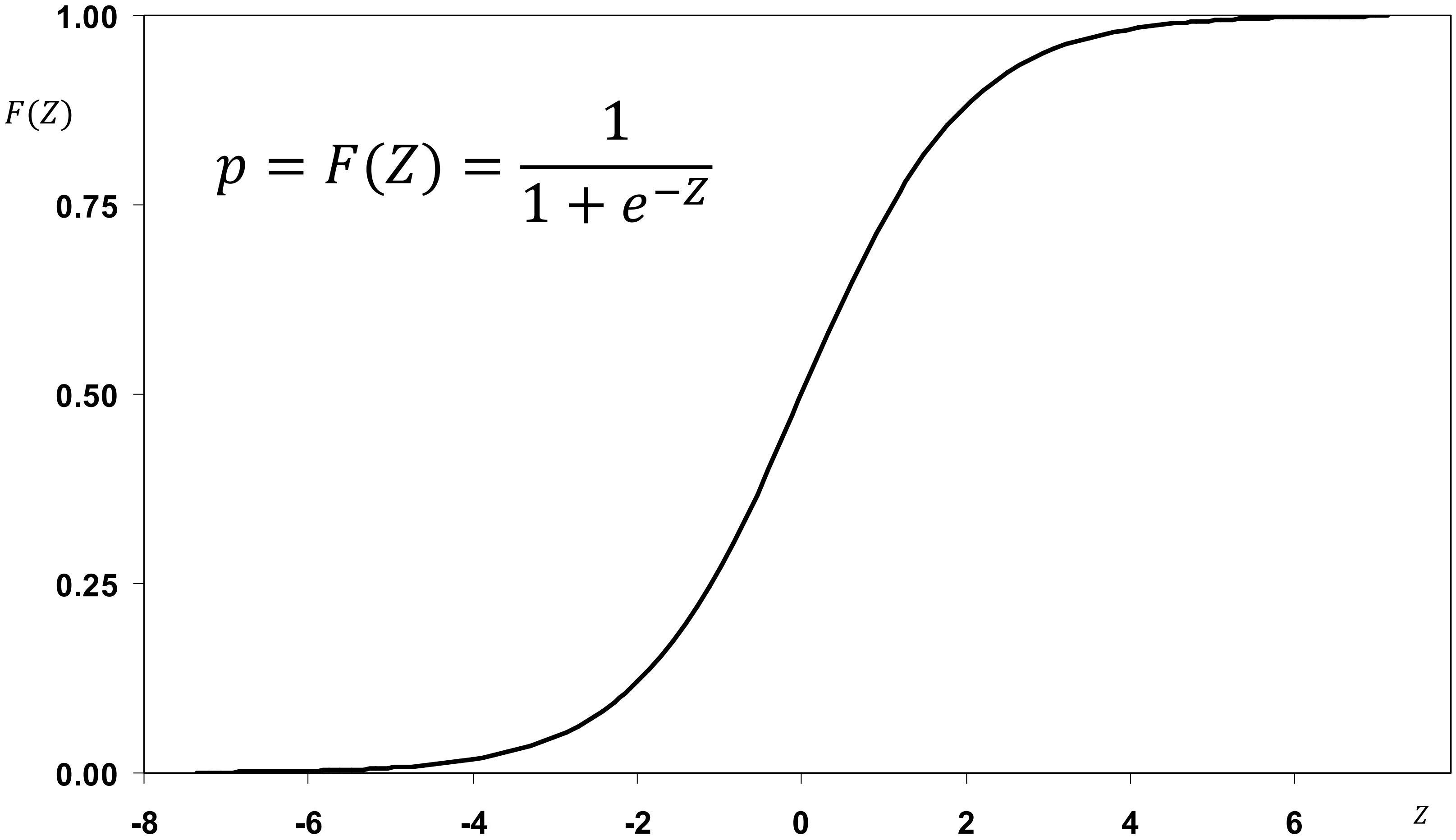
Тогда $P(Y_i = 1) = P(Y_i^* \geq 0) = P(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i \geq 0) =$

$P(\varepsilon_i \geq -\beta_0 - \beta_1 X_{1i} - \ldots - \beta_k X_{ki}) = P(\varepsilon_i \leq \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki}) =$

$= F((\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})/\sigma),$

что с точностью до нормировки совпадает с (*).

Логит - модель



Если функция F является логистической (формула для этой функции приведена выше), то соответствующая модель называется логит - моделью.

Логит - модель

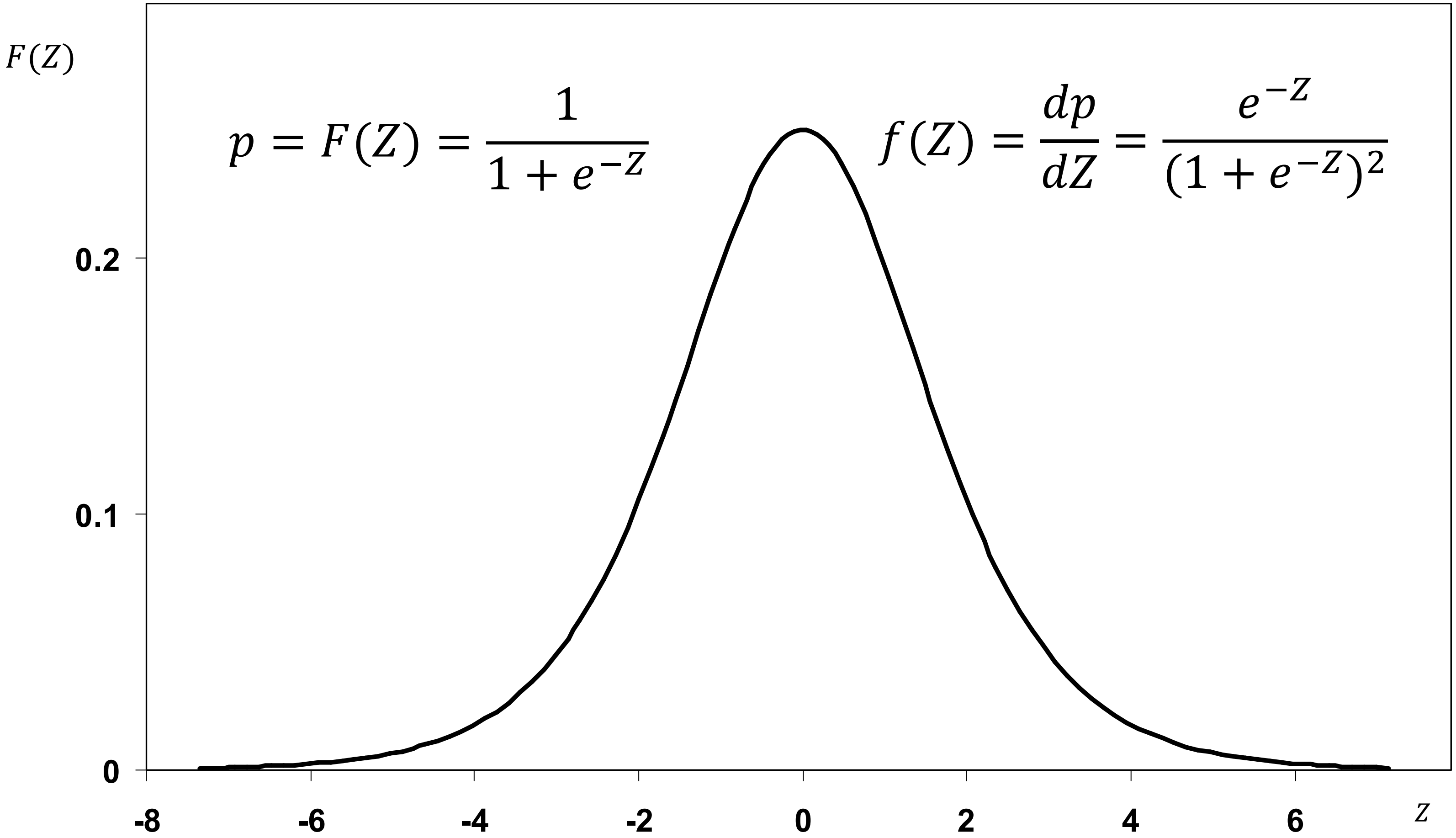
$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

$$\frac{dp}{dZ} = \frac{(1 + e^{-Z}) \times 0 - 1 \times (-e^{-Z})}{(1 + e^{-Z})^2}$$

$$= \frac{e^{-Z}}{(1 + e^{-Z})^2}$$

Производная функции F(*Z*) называется функцией плотности. Выше вычислена функция плотности для логистической функции.

Логит - модель



На рисунке изображен график функции плотности *f*(*Z*) для логистической функции.

Логит - модель

Для оценки параметров моделей бинарного выбора используется ММП.

Функция правдоподобия в этом случае имеет вид:

$$L(\beta) = \prod_{Y_i=1} F\,(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \prod_{Y_i=0} (1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})) =$$

$$= \prod_{i=1} [F(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})]^{Y_i} [1 - F(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})]^{1-Y_i}$$

Логит - модель

Логарифмическая функция правдоподобия:

$$l(\beta) = \ln L\,(\beta) = \sum_{i=1}^{n} [Y_i \ln F\,(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) +$$

$$+(1 - Y_i)\ln(\,1 - F(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}))]$$

Логит - модель

Необходимые условия экстремума первого порядка:

$$\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} X_{ji} \left[\begin{array}{c} \dfrac{f(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})}{F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})} Y_i + \\ + (1 - Y_i) \dfrac{(-f(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}))}{1 - F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})} \end{array} \right] = 0,$$

$$j = 0, \dots, k$$

Логит - модель

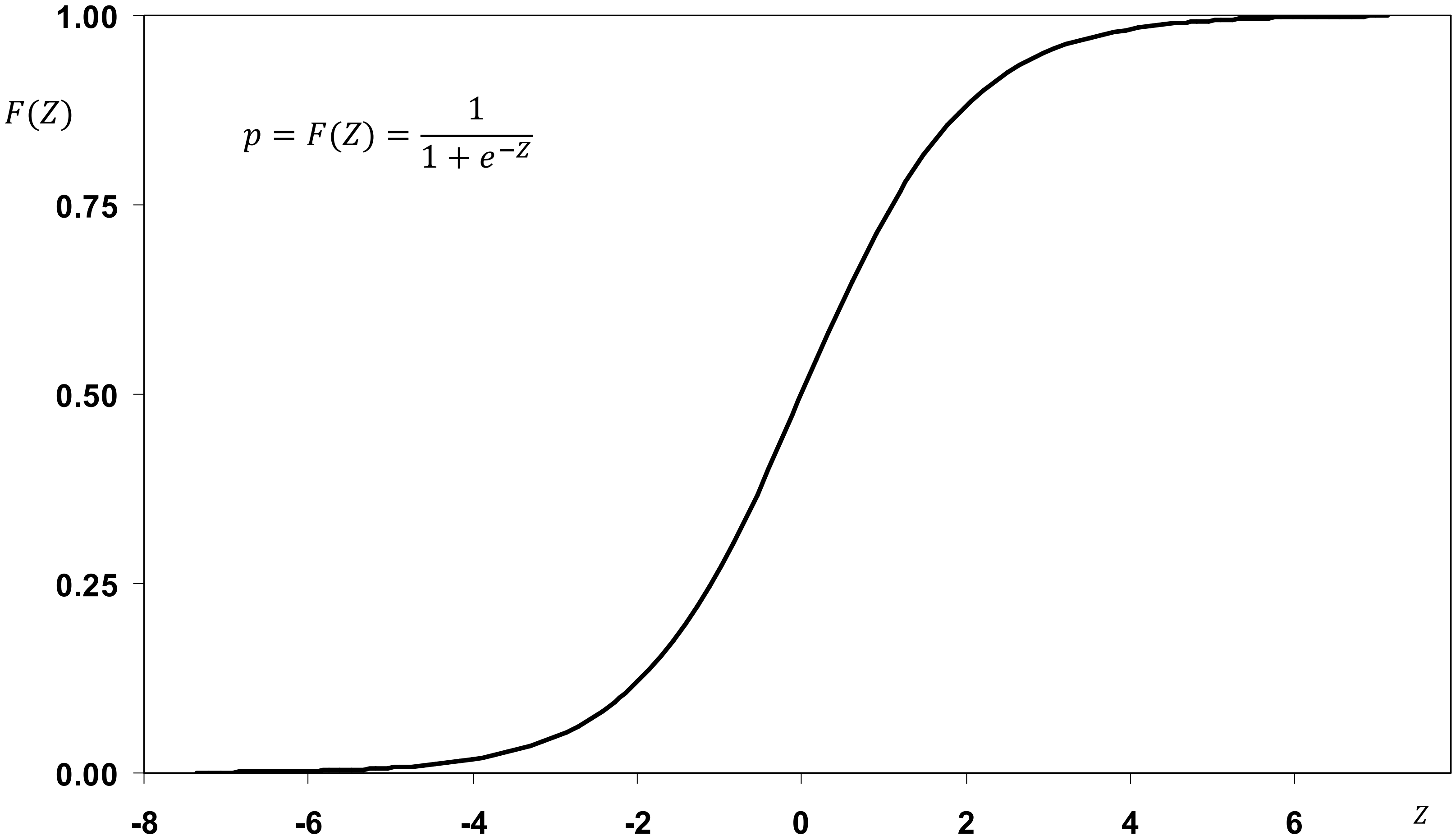
Для логит-модели эта система уравнений может быть сведена к виду :

$$\sum_{i=1}^{n} [Y_i - \Lambda(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})] X_{ji} = 0 \qquad j = 0, \ldots, k$$

Эта система уравнений решается с помощью численных методов.

Матрица Гессе для логит-модели является отрицательно определенной (детали см. Green, издание 7, с.691-692), поэтому найденное решение является максимумом.

Логит - модель



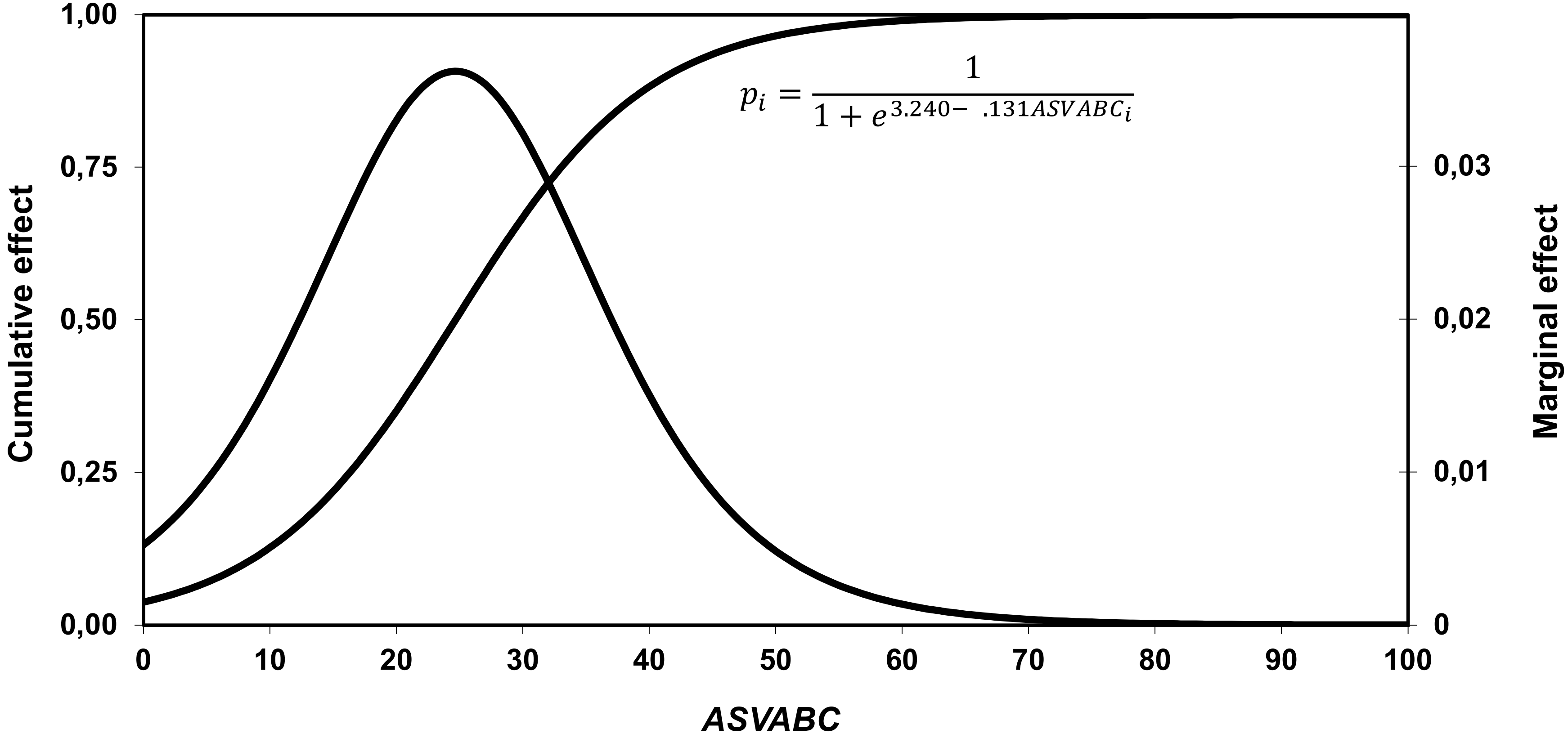
Пример использования логит – модели для оценки вероятности окончания средней школы. В качестве объясняющей выбрана переменная *ASVABC*.

Логит - модель

```
. logit GRAD ASVABC

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -104.45292
Iteration 2:   log likelihood = -97.135677
Iteration 3:   log likelihood = -96.887294
Iteration 4:   log likelihood = -96.886017

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428      5.86    0.000      .0874045    .1753206
       _cons |  -3.240218   .9444844     -3.43    0.001     -5.091373   -1.389063
------------------------------------------------------------------------------
```

Результаты оценивания.

Логит - модель



Оцененная модель.

Логит - модель

```
. logit GRAD ASVABC

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -104.45292
Iteration 2:   log likelihood = -97.135677
Iteration 3:   log likelihood = -96.887294
Iteration 4:   log likelihood = -96.886017

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428      5.86   0.000      .0874045    .1753206
       _cons |  -3.240218   .9444844     -3.43   0.001     -5.091373   -1.389063
------------------------------------------------------------------------------
```

$$\hat{Z} = -3.240 + 0.131 ASVABC$$

Коэффициент перед переменной ASVABC является значимым.

Логит - модель

$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

$$Z = \beta_0 + \beta_1 X_1 + \ldots \beta_k X_k$$

В случае нелинейных моделей говорят о предельном эффекте объясняющего фактора.

Логит - модель

$$p = F(Z) = \frac{1}{1 + e^{-Z}}$$

$$Z = \beta_0 + \beta_1 X_1 + \ldots \beta_k X_k$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(Z)\beta_i = \frac{e^{-Z}}{(1 + e^{-Z})^2}\beta_i$$

Предельный эффект объясняющего фактора X_i (если X – непрерывная переменная) – это частная производная по этой переменной. Вычисляется эта производная по правилу вычисления производной сложной функции. Предельный эффект i – го объясняющего фактора не является константой, а зависит от других переменных

Логит - модель

$$p(X_1, X_2, \ldots, X_j = 1, \ldots, X_k) -$$

$$p(X_1, X_2, \ldots, X_j = 0, \ldots, X_k)$$

Предельный эффект объясняющего фактора X_j (если X_j – dummy переменная).

Логит - модель

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963
```

```
Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428     5.86    0.000     .0874045    .1753206
       _cons |  -3.240218   .9444844    -3.43    0.001    -5.091373   -1.389063
------------------------------------------------------------------------------
```

В рассмотренном примере средний результат A*SVABC* равен 51.36.

Логит - модель

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963
```

$$Z = -3.240 + 0.131 \times 51.36 = 3.507$$

```
Logit estimates                                   Number of obs   =        540
                                                  LR chi2(1)      =      43.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.886017                       Pseudo R2       =     0.1836

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1313626    .022428      5.86    0.000     .0874045    .1753206
       _cons |  -3.240218   .9444844     -3.43    0.001    -5.091373   -1.389063
------------------------------------------------------------------------------
```

В этой точке *Z* равно 3.507.

Логит - модель

```
. sum GRAD ASVABC

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        GRAD |       540    .9425926    .2328351          0          1
      ASVABC |       540    51.36271    9.567646   25.45931   66.07963
```

$$e^{-Z} = e^{-3.507} = 0.030$$

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1+e^{-Z})^2} = \frac{0.030}{(1+0.030)^2} = 0.028$$

e^{-Z} равно 0.030. Следовательно, *f*(*Z*) равно 0.028.

Логит - модель

```
. sum GRAD ASVABC
```

Variable	Obs	Mean	Std. Dev.	Min	Max
GRAD	540	.9425926	.2328351	0	1
ASVABC	540	51.36271	9.567646	25.45931	66.07963

$$Z = \beta_0 + \beta_1 \bar{X} = -3.240 + 0.131 \times 51.36 = 3.507$$

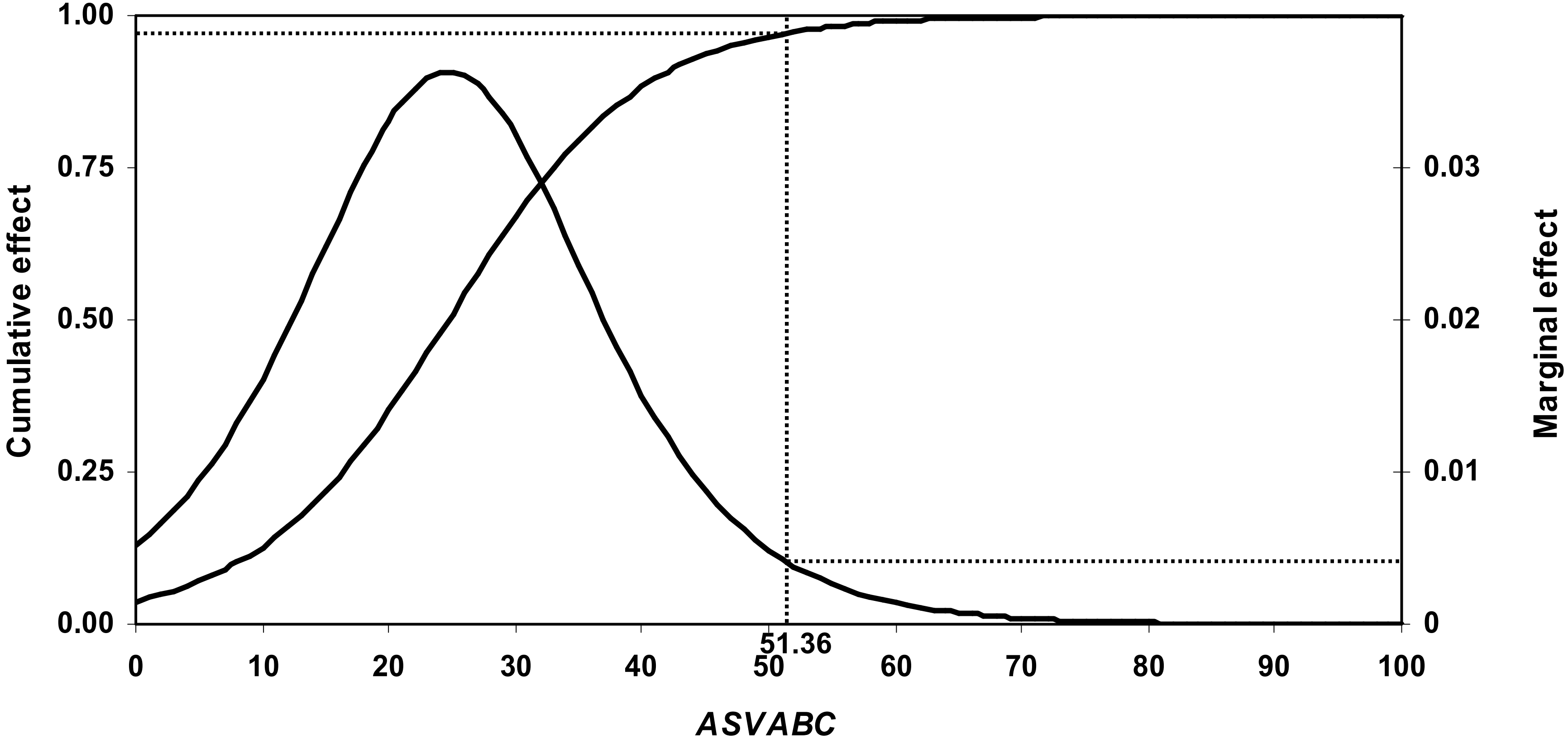
$$e^{-Z} = e^{-3.507} = 0.030$$

$$f(Z) = \frac{dp}{dZ} = \frac{e^{-Z}}{(1+e^{-Z})^2} = \frac{0.030}{(1+0.030)^2} = 0.028$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(Z)\beta_i = 0.028 \times 0.131 = 0.004$$

Предельный эффект для имеющего средний результат тестирования равен 0.004. Это означает, что при увеличении результата тестирования *ASVABC* на 1 балл вероятность закончить школу возрастает на 0.4 процента.

Логит - модель

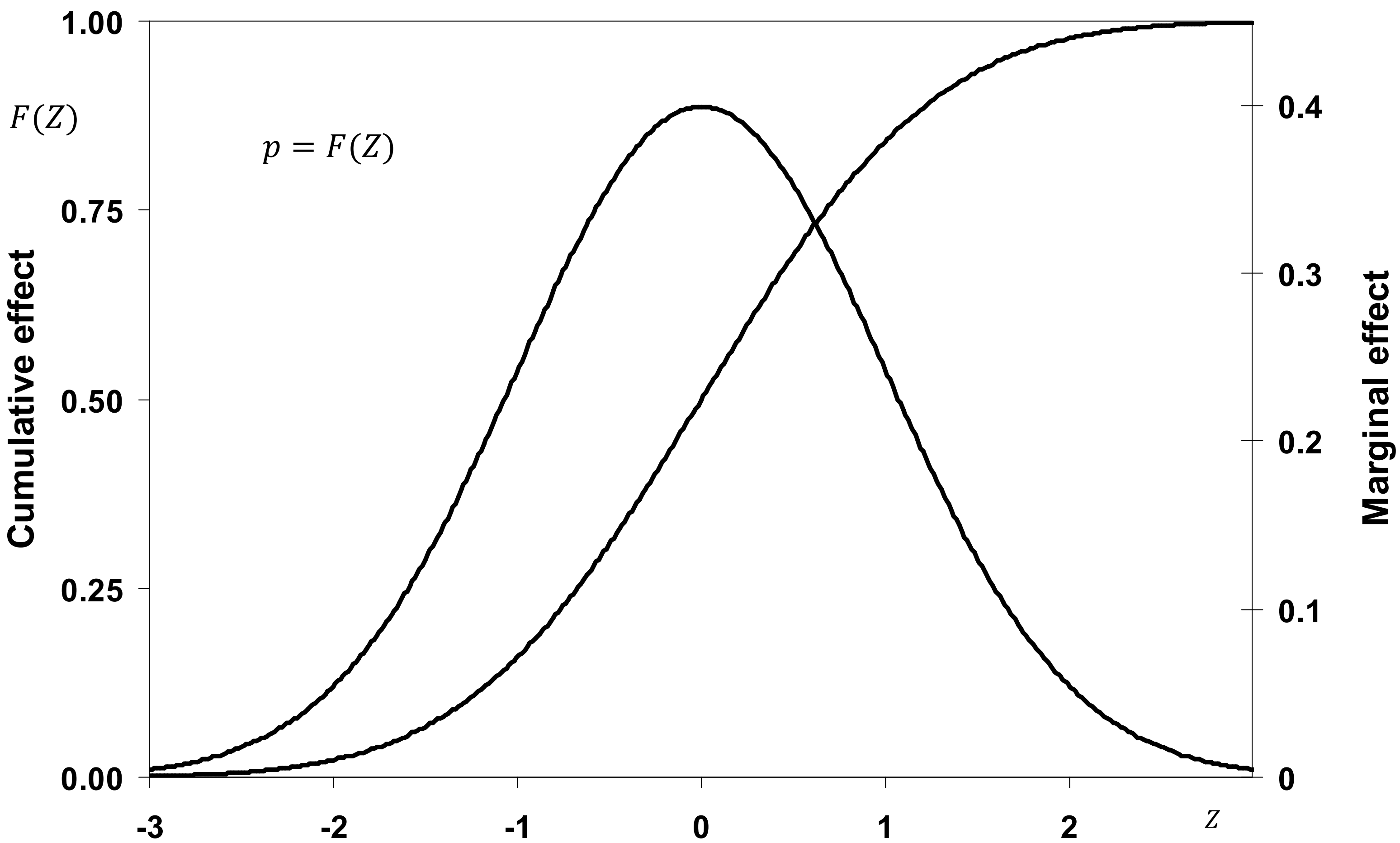


Предельный эффект при среднем результате очень мал. Это связано с тем, что вероятность закончить школу при средних результатах и так очень велика.

Логит - модель

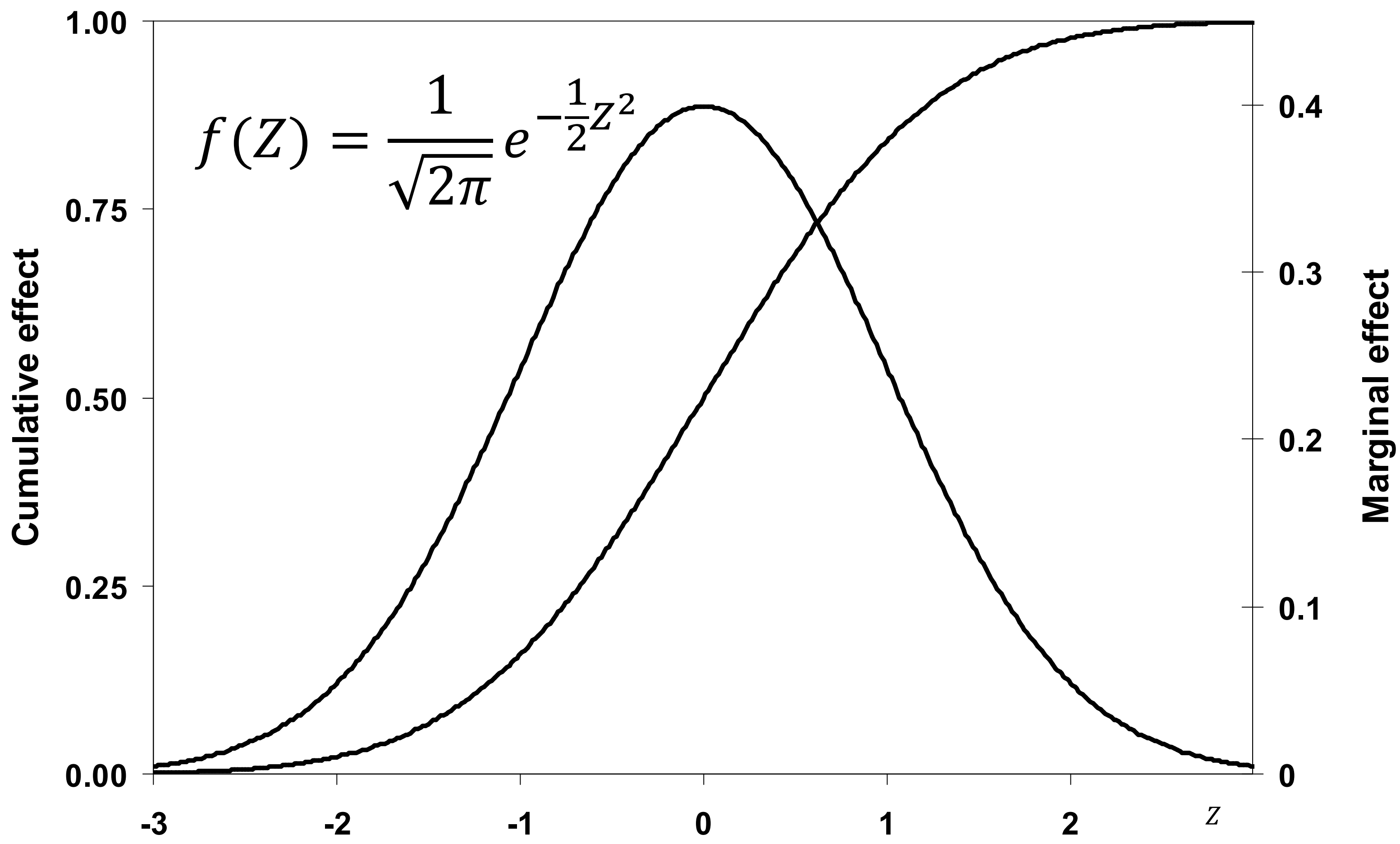
В пакете STATA предельные эффекты объясняющих переменных можно получить с помощью команды mfx

Пробит - модель



Для пробит – модели в качестве S – функции выбирается функция распределения стандартного нормального распределения.

Пробит - модель



Выше приведена функция плотности. Оценки коэффициентов находятся по методу максимального правдоподобия.

Пробит - модель

```
. probit GRAD ASVABC SM SF MALE

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -98.195303
Iteration 2:   log likelihood = -96.666096
Iteration 3:   log likelihood = -96.624979
Iteration 4:   log likelihood = -96.624926

Probit estimates                                  Number of obs   =        540
                                                  LR chi2(4)      =      44.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.624926                       Pseudo R2       =     0.1858

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .0648442   .0120378     5.39   0.000      .0412505    .0884379
          SM |  -.0081163   .0440399    -0.18   0.854      -.094433    .0782004
          SF |   .0056041   .0359557     0.16   0.876     -.0648677    .0760759
        MALE |   .0630588   .1988279     0.32   0.751     -.3266368    .4527544
       _cons |  -1.450787   .5470608    -2.65   0.008     -2.523006   -.3785673
------------------------------------------------------------------------------
```

Результаты оценки пробит- модели.

Пробит - модель

```
. probit GRAD ASVABC SM SF MALE

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -98.195303
Iteration 2:   log likelihood = -96.666096
Iteration 3:   log likelihood = -96.624979
Iteration 4:   log likelihood = -96.624926

Probit estimates                                  Number of obs   =        540
                                                  LR chi2(4)      =      44.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.624926                       Pseudo R2       =     0.1858

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .0648442   .0120378     5.39   0.000     .0412505    .0884379
          SM |  -.0081163   .0440399    -0.18   0.854     -.094433    .0782004
          SF |   .0056041   .0359557     0.16   0.876    -.0648677    .0760759
        MALE |   .0630588   .1988279     0.32   0.751    -.3266368    .4527544
       _cons |  -1.450787   .5470608    -2.65   0.008    -2.523006   -.3785673
------------------------------------------------------------------------------
```

Как и для логит – модели, не существует интерпретации полученных оценок коэффициентов. С их помощью можно рассчитать предельные эффекты.

Пробит - модель

$$p = F(Z)$$

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(Z)\beta_i = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}\right)\beta_i$$

Напомним, что предельный эффект объясняющего фактора X_i рассчитывается как частная производная от *X_i*. Предельный эффект i – го объясняющего фактора не является константой, а зависит от других переменных

Пробит - модель

```
. sum GRAD ASVABC SM SF MALE
```

Variable	Obs	Mean	Std. Dev.	Min	Max
GRAD	540	.9425926	.2328351	0	1
ASVABC	540	51.36271	9.567646	25.45931	66.07963
SM	540	11.57963	2.816456	0	20
SF	540	11.83704	3.53715	0	20
MALE	540	.5	.5004636	0	1

Таблица дескриптивных статистик для переменных.

Пробит - модель

Probit: Marginal Effects

	mean	*b*	product	*f*(*Z*)	*f*(*Z*)*b*
ASVABC	51.36	0.065	3.328	0.068	0.004
SM	11.58	–0.008	–0.094	0.068	–0.001
SF	11.84	0.006	0.066	0.068	0.000
MALE	0.50	0.063	0.032	0.068	0.004
constant	1.00	–1.451	–1.451		
Total			1.881		

$$\frac{\partial p}{\partial X_i} = \frac{dp}{dZ}\frac{\partial Z}{\partial X_i} = f(Z)\beta_i$$

Оцениваем предельные эффекты для объясняющих факторов.

Odd Ratio

Для логит-модели

$$OR = \frac{\Pr(Y = 1)}{\Pr(Y = 0)}$$

Отношение вероятности «удачи» и «неудачи»

$$\ln(OR) = \beta_0 + \beta_1 X_1 + \ldots \beta_k X_k$$

Если Xj изменится на 1 то OR изменится в $\exp(\beta_j)$

раз.

Показатели качества оценки моделей бинарного выбора

R^2 МакФаддена (McFadden) определяется формулой:

$R_{MF}^2 = 1 - \frac{\hat{l}}{l_0}$, где $\hat{l}$ - логарифмическая функция правдоподобия для модели (11.2) в точке максимума, l_0 - максимум логарифмической функция правдоподобия для модели, в которую включена только константа,

псевдо R^2 (Amemiya, 1981):

$$PseudoR^2 = 1 - \frac{1}{1+\frac{2}{n}(\hat{l}-l_0)}.$$

Показатели качества оценки моделей бинарного выбора

Альтернативный критерий оценки качества модели является сравнение точности прогнозирования по оцененной модели бинарного выбора.

Прогноз по модели бинарного выбора (11.2) обычно строится следующим образом:

$$\begin{cases} \hat{Y}_i = 1, & если \quad F(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}) > 0.5, \\ \hat{Y}_i = 0, & если \quad F(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}) \le 0.5 \end{cases},$$

что, с учетом симметричности функций плотности относительно нуля равносильно

$$\begin{cases} \hat{Y}_i = 1, & если \quad \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki} > 0, \\ \hat{Y}_i = 0, & если \quad \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki} \le 0. \end{cases}$$

5

Показатели качества оценки моделей бинарного выбора

Доля неверных прогнозов вычисляется по формуле:

$$wr_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 .$$

Эта доля неверных прогнозов интересна в сравнении с долей неверных прогнозов самой простой бинарной модели, в которую включена только константа. Для этой просто модели прогнозы строятся совсем просто:

$$\begin{aligned} &\hat{Y}_i = 1, \quad \textit{если} \quad \hat{p} > 0.5, \quad \textit{где} \quad \hat{p} - \textit{доля} \quad \textit{единиц} \quad \textit{в} \quad \textit{выборке}, \\ &\hat{Y}_i = 0, \quad \textit{если} \quad \hat{p} \le 0.5. \end{aligned}$$

Доля неверных прогнозов для простой модели вычисляется так:

$$\begin{cases} wr_0 = 1 - \hat{p}, \quad \textit{если} \quad \hat{p} > 0.5, \\ wr_0 = \hat{p}, \quad \textit{если} \quad \hat{p} \le 0.5. \end{cases}$$

Отметим, что пороговое значение 0.5 иногда изменяют, чтобы достичь более высоко точности предсказания.

Показатель качества подгонки модели рассчитывается по формуле:

$$R_p^2 = 1 - \frac{wr_1}{wr_0} .$$

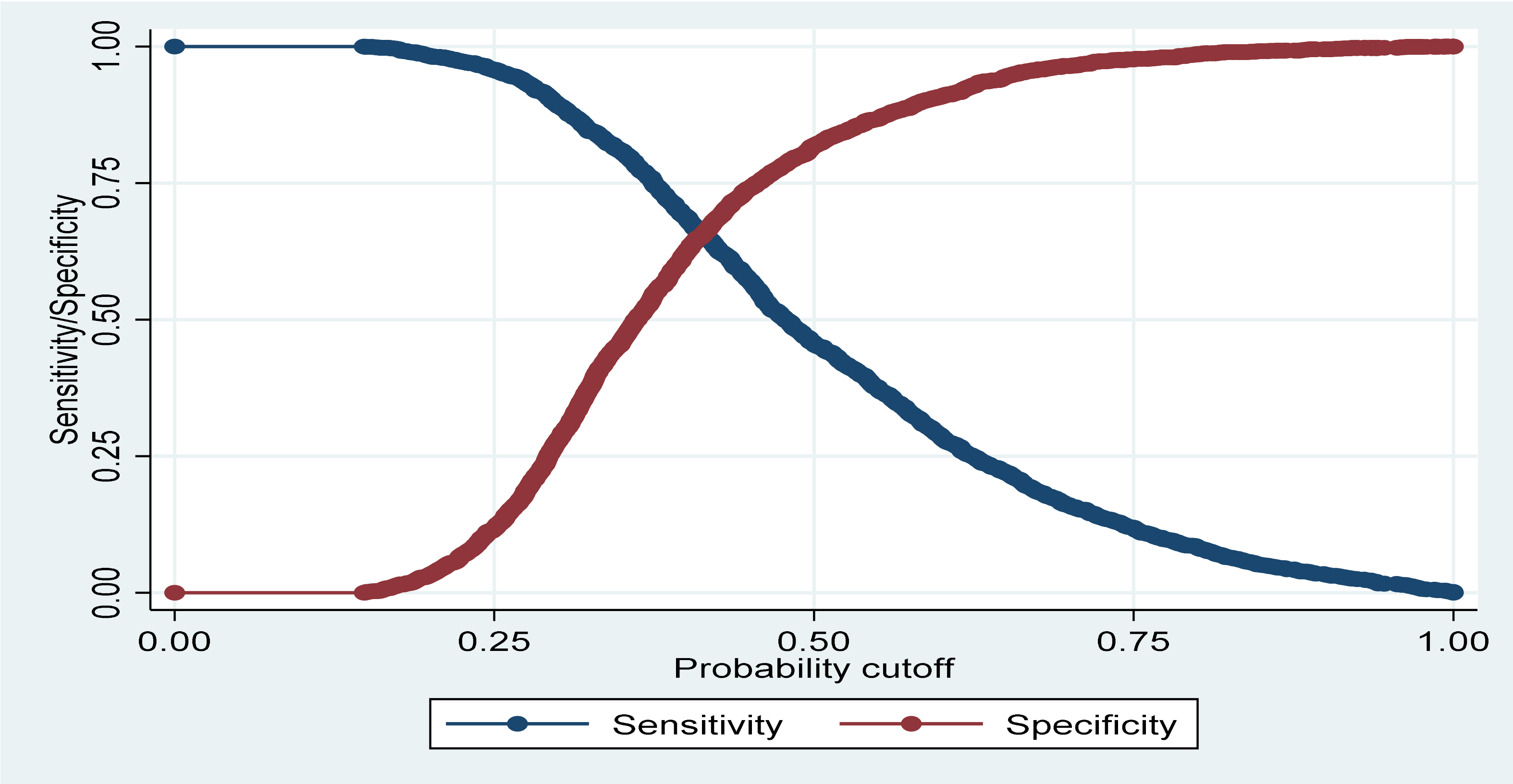
Показатели качества оценки моделей бинарного выбора

	True		
Classified	D	~D	Total
+	793	397	1190
-	943	1789	2732
Total	1736	2186	3922

Classified + if predicted Pr(D) >= .5
True D defined as male != 0

Sensitivity	Pr(+| D)	45.68%
Specificity	Pr(-|~D)	81.84%
Positive predictive value	Pr(D| +)	66.64%
Negative predictive value	Pr(~D| -)	65.48%
False + rate for true ~D	Pr(+|~D)	18.16%
False - rate for true D	Pr(-| D)	54.32%
False + rate for classified +	Pr(~D| +)	33.36%
False - rate for classified -	Pr(D| -)	34.52%
Correctly classified		65.83%

Показатели качества оценки моделей бинарного выбора

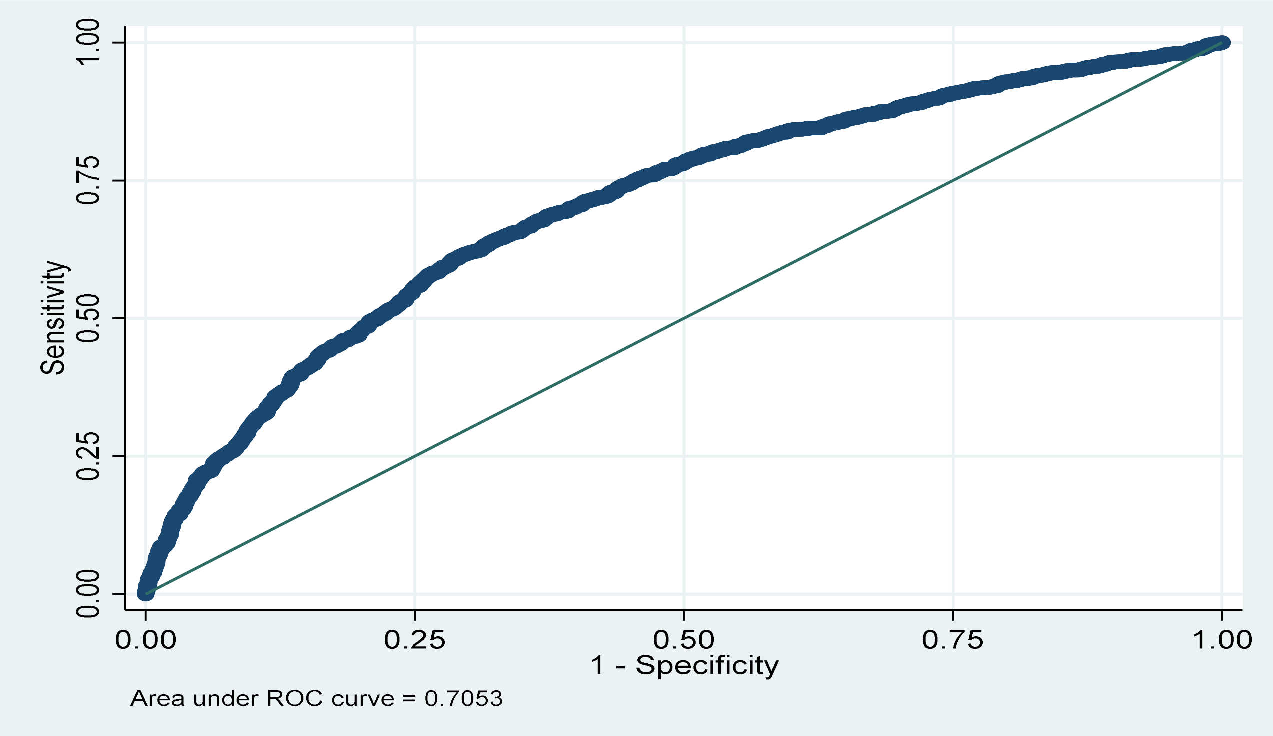


Sensitivity – доля правильно идентифицированных 1,

Specifity – доля правильно идентифицированных 0.

ROC - кривая

ROC – received operating characteristics



AUC = 0.5 – модель не лучше случайного гадания,

AUC < 0.5 – модель хуже случайного гадания